AD NUMBER

AD002148

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution: No Foreign

AUTHORITY

AFCRL ltr 30 April 1968

AD – 2148

UNCLASSIFIED

HFORL Reprint Report 2

Human Factors Operations Research Laboratories
Air Research and Development Command, USAF
Bolling Air Force Base, Washington 25, D.C.

Reprinted from The Journal of the Acoustical Society of America, Vol. 24, No. 6, 595–596, November, 1952

Control Tower Language

F. C. Frick and W. H. Sumby
Human Resources Research Laboratories, Bolling Air Force Base, Washington, D. C.
(Received July 11, 1952)

Shannon and others have estimated that written English is about 50 percent redundant. These estimates are arrived at by considering linguistic constraints on our use of speech symbols; they do not consider additional restrictions imposed by the audience and the situation in which the speaker finds himself. In order to estimate the effects of such nonlinguistic constraints, an informational analysis has been made of the "sublanguage" used in the control of aircraft by Air Force control tower operators. When the situational, as well as linguistic, contexts are taken into account, the estimated redundancy is raised to 96 percent.

HOW informative a given speech sample is—how much we say—depends to a large extent on how much we might have said.

"Speaking English" imposes certain constraints on our use of linguistic symbols. The individual elements—phonemes, letters, words—which make up our vocabulary are not used equally often and they are not combined at random. In addition, the rules of grammer and the desire to make sense impose restrictions on the order in which groups of these symbols are strung together. In short, knowing the language of the speaker implies that we know quite a lot about the rules which govern his selection of speech symbols. It means that we are less uncertain about what will be said than we would be if the speaker did not operate under these known restrictions.

If we define information transmission as the rate of change in our *a priori* uncertainty, then we must say that the English language conveys less information than it could convey. The reduction in information that occurs when we pass from what might have been said (given the same vocabulary) to what actually is said is called "redundancy."

Special languages—cant, patois, technical talk—all involve additional restrictions and, for the listener familiar with them, entail increased redundancy. What we wish to do here is to consider one particular sublanguage of English and estimate the reduction in information transmission from what could be transmitted using alphabetical sequences. The language considered is that used in the control of aircraft by the operator of a control tower at a military airbase. It is made up of sentences like: "Air Force 5264. Ready number one in take-off position. Over." Or: "Extend your base. We have a C-54 on final."

These messages form a subset of the set of all possible English sentences—which in turn is a subset of all possible English letter sequences.

What this means can be illustrated fairly simply. All that is required is for you to imagine the set of all possible permutations of the English alphabet and a space symbol. Each point in this set represents a possible sequence of letters.

We could now, in principle, go through this set of letter sequences and select out those sequences that are acceptable as English sentences. This will clearly cut down the size of our original set. The knowledge that we are operating within this subset of English sentences reduces our initial uncertainty. And Shannon[1] has developed a technique that permits us to estimate how much this reduction amounts to. Essentially, we ask our subjects to guess at successive letters of English text. When we do this, we find that the subject can exploit his implicit knowledge of the statistical structure of the language and predict the English sequence considerably better than he can predict sequences of letters chosen at random.

If we now ask the same subjects to guess at the text of control tower messages, we find that prediction is still farther improved. The uncertainty of our subjects was, on the average, about 28 percent of their uncertainty regarding random sequences of letters and spaces—that is, of course, subjects who are familiar with control tower language. They were, in fact, control tower operators. (Actually, people who are not practiced in this strange sublanguage do just about as well. The airplane and its control has apparently been absorbed into our linguistic culture.)

However, this is still an overestimate of the uncertainty, or unpredictability, of messages in the actual situation. The figure above is arrived at when we select letter sequences at random from the set of admissible tower messages. In any given instance, the message is, of course, not generated at random. The pilot, for example, knows whether he is landing or taking off and this (the situational context) further restricts the set of possible messages.

In order to estimate these situational constraints we described a group of hypothetical situations to [illegible] Air Force pilots and [illegible] them to predict the tower message. For practical reasons we could not use Shannon's guessing game technique, so we adopted another device.

The messages the pilots predicted for us are rather easily split up into content units—phrases that have the same meaning—e.g., gear down and locked, gear in the green—or that differ only with respect to the numbers or place names involved—runway zero nine, runway

[1] C. E. Shannon, Bell System Tech. J. 30, 50-64 (1951).

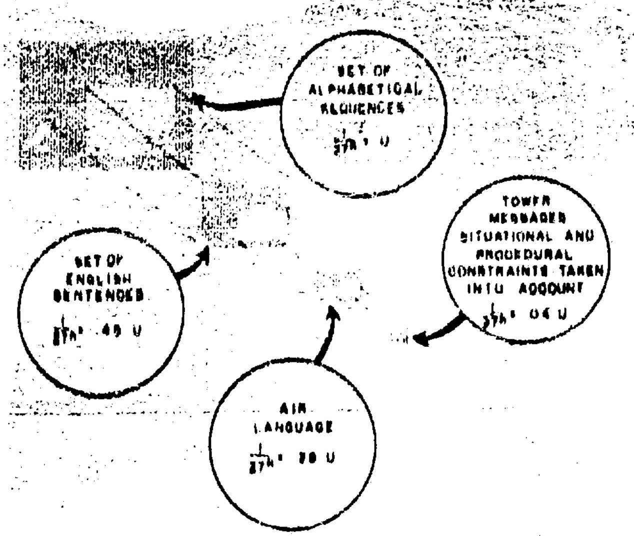

FIG. 1. Schematic representation of the reduction in size of the message set and consequent reduction in uncertainty (U) as additional message restrictions are taken into account.

two zero. A message is made up of a sequence of these message units. And it is at this level that the immediate situation seems to operate by determining what message units will be selected and in what order they will occur. These units, thought of as letter sequences, are subject to the linguistic constraints that we have already estimated. In effect, we have the physical situation determining the gross uncertainty of the message (in terms of content units) and the linguistic constraints determining the uncertainty of the units (as sequences of letters and spaces).

To illustrate, let us consider a specific case. The pilot is told: *You are coming in to land. Ceiling and visibility unlimited. You have just called in: "Andrews Tower. This is Air Force 1234. Eight miles south of your station. Landing instructions please."* Each statement sets up additional restrictions on the set of possible tower messages.

If we now look at the actual distribution of predicted messages which we obtained in this case, we find that they are sequences of selections from only 13 message units. Furthermore, these elements do not appear with equal frequency, nor does any single message include all 13 elements. A large number of possible messages thus turn out to be impossible—or at any rate highly improbable. The message set is thus further reduced.

Lastly, this language has, at the level of content units, its own peculiar grammer, known as *RT procedure*. The pilot knows this procedure (though ours didn't know it very well) and this procedure fixes the order in which particular units are selected to make up the predicted message.

With such a small sample we cannot determine the total effect of these sequential dependencies in restricting the message class. But we can estimate these constraints over two or three units. When we do this for this particular case, we find that the pilot's uncertainty of the message (in terms of content units) is about 80 percent less than it would have been if the units were equally likely and messages were generated by a random selection of content units (with no repeats).

This figure of 80 percent varies somewhat with the particular situation. For the 15 different situations that we chose, it averages 80 percent.

In other words, the situational context, including the pilot's knowledge of procedures, reduces the gross uncertainty of the message about 80 percent. Linguistic constraints reduce the residual uncertainty, in terms of letter sequences, another 72 percent—giving us an estimated redundancy, with respect to what could have been conveyed using letter and space sequences, of approximately 96 percent. The entire process is illustrated in Fig. 1.

This is a large figure and any communication system which tolerates so much redundancy is in some sense inefficient. The situation is not as bad, however, as might at first be imagined. This is a noisy system. In part, at least, it is unavoidably noisy. It is also a system with a low tolerance for error—planes are more expensive than transmission time or band width—and redundancy is an effective means of combating noise and error. In other words, we may not transmit much information, but what is transmitted is important and the high degree of redundancy is a form of insurance—a sort of running check on our transmission.

AD-

2148

UNCLASSIFIED